IEEE International Conference on Advances in Engineering & Technology Research (ICAETR - 2014), August 01-02, 2014, Dr. Virendra Swarup Group of Institutions, Unnao, India

Bridging the Gap between Disabled People and New Technology in Interactive Web Application with the Help of Voice

Dilip Kumar
Faculty of Computer Science & Engineering
Shri Ramswaroop Memorial University
Lucknow, Uttar Pradesh - 225003, India.
dilipkumarjnu@gmail.com

Abhishek Sachan
Faculty of Computer Science & Engineering
Shri Ramswaroop Memorial University
Lucknow, Uttar Pradesh - 225003, India.
sachan.abhishek001@gmail.com

Abstract—**Major area of developments in the field of voice recognition is more natural way of interacting with web technology have been taking place. The concept that computers can comprehend our various gestures by eyes, voices, touch and our different movements to interact is called the Easy User Interface. Today, many elements are available in mobile phones, PCs and in other devices but the lack of voice to voice application on web are not efficient to the disabled people. Speech technologies particularly play a vital role in this evolving technology. Significant advancement in automatic voice recognition for well defined applications like dictation and medium vocabulary transaction processing assignments in web application is comparatively controlled environments have been made on web server. But, automatic speech recognition still has to reach a level needed for speech to become a completely pervasive user interface because even in clean acoustic surroundings in web application. Visual voice recognition, however, is a promising source of extra voice information and it has successfully exhibited to enhance noise robustness of automatic voice recognizers, so thereby promising to enhance their usability in the web application by use of voice interaction in web. In this paper, the main objective is to provide more easy interaction between user (disabled people) and web application. Today we are using web text by typing the text and in early days Google chrome enhances their functionality adding voice search. But we are trying to accessing all web text with the help of voice. We are using JSAPI components of speech recognition from the java technology, i.e., the voice and the text components along with the latest advancements made in the field of web uses. Further, the research goes beyond the recent advancements and discusses the future scope of speech recognition technology and mentions some likely future developments, evaluating each on the basis of its performance.**

Keywords- speech recognition; web speech recognition technology; CRD; pronounced; human computer web interaction; special education; teaching material.

I. INTRODUCTION

In modern era where voice and visual technologies are used in widely extended. Voder provide the first speech synthesizer. This application is easily operated on the web server. The sound level is quieted well. In past days machine talk is the way through which human interact but now a day's machine and human interaction is the main motto, which give the invention of speech recognition. This speech recognition concludes many areas like mathematics, artificial intelligence, machine learning, statics and other electronics devices (microphone, processor crd, sound card technology)[1,2,3]. This type of developed application is capable to understand specific context of sentences, words, commands and makes the user flexible to input as a voice and also help to control all the web application text available on the web server as well as web reading. Many implementations is already done on English speech recognition, but creates problem with the person with disability, in this paper main focus on the speech reorganization and get the result in voice too.

To easily communicate with the people voice is the best medium, not only peoples but also to the web it is more flexible way, from early days for man machine interaction voice is the best medium to interact for communicating the computer with more humanly nature like voice to text and text to voice conversion have been studied. To make this communication should more flexible. This is done by transmitting voice not only with ideas, concepts but also with information carrying attitude, emotion and individuality of speaker.

Hinnselwood in 1917 introduced the term "congenital word blindness" to describe this disability. Strauss and Warner in 1941 focus to the cases of the person with sufficient intellects are unsuccessful at the school due to experiencing reading difficulty. This type of difficulty is also define by the national committee like "national joint committee on learning disability" in 1988 emphasis on learning disability, basically learning disability refers to the faction of disorders manifested by important problem in the attainment and use of speaking, writing, speaking, reasoning mathematical disability on the web[4].

For improving the mental process of speech of students, asked to focus on interpretation, relating, classification, these all term may causes disorders in the brain like classification is the consistent arrangement of specific items or things based on certain categories. So from the mental aspect of speech of students should aware of classification of words, objects subjects, sentences, animals, pants, facts, events etc. in the terms of web.

A. What is Speech Recognition

To modify source speaker in to the target speaker voice conversion (vc) system plays an important role. Basically speech signal provides different types of information, different fields of speech technology focuses on different information. The main focus of voice conversion is speaker identity. Voice Conversion works on two major problem deals with speech. Firstly, characteristics identification of speaker during analysis phase and secondly in synthesis phase where replacement of source characteristics with the target characteristics. These operations are independent from each other.

B. Objective of Study

It is an era of Internet that have responsibility where individuals use, follow and web development technology for better communication language is the key for personnel and cultural development [5]. Due to the increasing use of audio visual communication tools invites people to use new modernized tools in education. For the success of students is to adapt the modernized educational equipment by voice. Multi tools are always beneficial then single tool for educational purposes in web application. This is a modernized period where the teaching based upon web technology and media is better than verbal teaching. It increases the level of learning, teaching and also provides solid information. Furthermore it improves students speaking and listening comprehension skills by the help this application [6].

C. Required Material & Techniques

In this paper we introduces the software and API that contribute to the problem regarding to the problem mentioned above like Bridging the Gap between Disabled People and New Technology in Interactive Web Application with the Help of Voice. This software contains many distinctive features like integrated speech reorganization as well as audio visually enriched boundary, which provides user better pronunciation activities and recognize the word.

II. RELATED WORK OF SPEECH RECOGNITION IN WEB APPLICATION

Speech awarding software is available in large quantities in the market. Many renewed software are developed by many prestigious company like IBM [11], Philips [12]. One of the interactive speech reorganization software developed by Microsoft is genie. Also various voice navigation applications and speech recognizer are developed like sphinx [7], sonic, voice [9][10].

In speech recognition process surroundings voices are the worst part, it confuses the recognizer to understand the real voice that are supposed to hear [8]. This type of voice recognizer is used in robots. One such recognizer has been states for medical robots [13] that, despite of the inevitable motor noises, makes it communicate with the people efficiently. It is made possible by using a noise-type-dependent acoustic model corresponding to a performing motion of voice in the web speech. The optimizations of speech recognition such as HP Smart Badge IV embedded system [15] has been proposed to reduce the energy consumption. While there maintaining the quality of the web application. Another such scalable system has been proposed in [14] for DSR (Distributed Speech recognition) by combining it with scalable compression and hence reducing the bandwidth requirement on the web server. Many capabilities of current speech recognizers in the field of human computer interaction are described in [16] like Voice Banking and Directory Assistance.

Speech recognition allows us to provide input the voice to an application with your voice. Just like clicking with your mouse, typing on your keyboard, or pressing a key on the phone keypad provides input to an application, speech recognition allows you to provide input by talking. In the world of internet, we need a microphone to be able to do this. In the Voice control world, all you need is a internet.

For example, if we might say something like "How are you", to which your voice to be converted into text and application replies by the help of voice on the server.

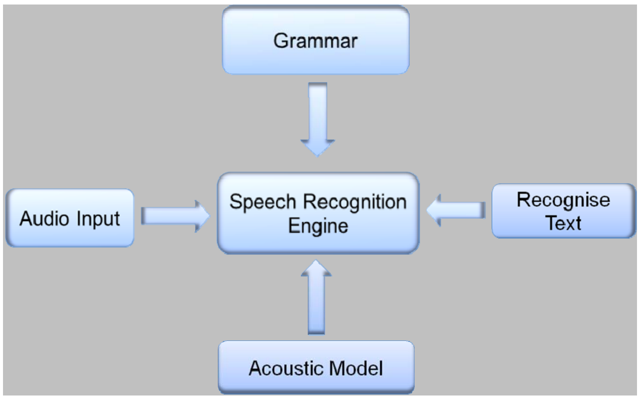

Figure.1. Process of Converting Speech into Text

Figure.1. shows speech reorganization engine recognized speech. The main function of this engine is to translate the speech into text so that application can understand it. The application basically does one of the two things:

- Does interruption of the result of recognizing of the speech. it acts in application as a command and control application.
- Appear as a dictation application

Some of the Fundamental of Speech Recognition:

A. Utterances

When the user tries to speak something is called utterance. It is steam of speech between the salience.

B. Pronunciations

Basically speech reorganization engine takes inputs of data, models, and algorithms to convert the text in to the speech. A particular piece of conformation that the process uses in the engine is called pronunciation.

C. Grammar

Speech reorganization engine works on certain domain is called grammar. In this all the utterances of speech is compared with the words and phrases in the active grammar.

Here are some voice reorganization programs are available:

- Windows 7: Most recent version of Microsoft contains this type of reorganization system. It provide the many application are controlled by the voice such as opening browser, opening and closing of paint and also other work being done.
- Dragon NaturallySpeaking: Dragon is the world's best-selling speech recognition software. It turns your talk into text and can make virtually any computer task easier and faster. From capturing ideas and creating documents, to email and searching the web, to using simple voice commands to control many of the popular programs you use every day at home, work – and beyond.
- Google Chrome browser: In the current system the Google chrome browser provide application of searching text into by the help of voice but it is limited to work that is only search in the form of text.

III. PROPOSED METHODOLOGY

Two main technologies required are as: speech reorganization and synthesis. Speech synthesis for commercial technology and speech reorganization is also supported by the academic and commercial systems, but in certain boundaries. Versatility and accuracy is the main trade off principle between the speech recognition. The desired speech technology it is impossible to investigate the real aspects of voice such as acceptance or satisfaction.

We have to be focused on realistic studies so that we can focus on the future developing of the applications. The crux problem related with the designing to determine the user acceptance without use of technology, so we decided to use a wizard of web speech API approach, with the human operator performing the speech reorganization.

The group meetings are occurring to focusing on services and interaction techniques. These groups contain 6-8 participants and 2 moderators. The participants give brief descriptions of the proposed applications. These participants also helps to determine the scenarios that how the application can be used. They portraying the users while other performing other different services. Some recording is also made for the future perspective. These meetings helps to develop a list of potential services and to get original glance of what interaction might look like for using of web.

There are large list of inputs like brainstorming, focus groups and literature survey. randomly we take four prototype us weather, headline news, messages and stock market results on the web application that input given by the voice and get the result with the help of voice.

The usability studies carried out with the help of prototype. These studies are helps to determine the usefulness of the provided services, common navigation path between and with i n the services. The usability studies are performed in to two stages: with small groups and with the large groups.

Both are performed in the same manner. Firstly a short description is given to the participants after that list of task is performed. A participant use natural voice language to complete assigned tasks by this user got the result to explore the system freely. Generated log file for the users contains time stamp's requests and replies, for later analysis of performance audio recording get stored on the server.

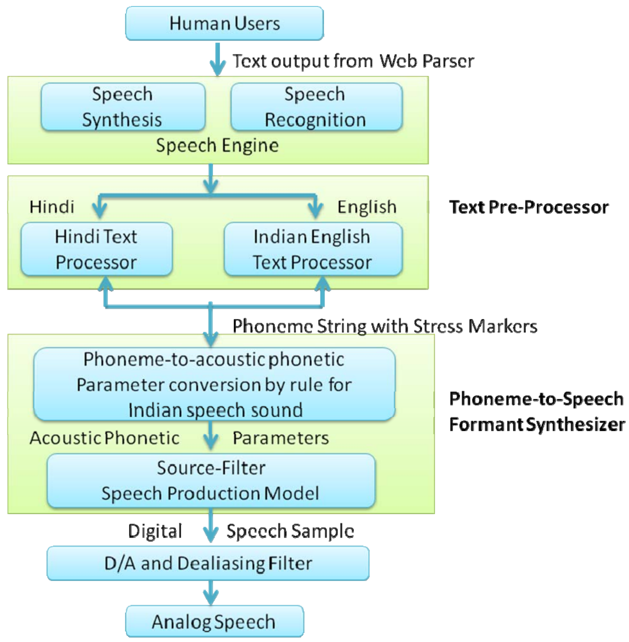

Figure.2.Proposed Model for Voice to Voice in Web application

IV. IMPLEMENTATION OF MODEL

Through web speech API speech synthesis and speech recognition adds up in the process. The post temporarily covers the last e.g. API recently added in Google chrome is x-webkit.

For supporting command and control reorganization dictation systems and speech synthesis the java speech API is used as an application programming interface.

Two core technology used in this proposed model are:

A. Speech Synthesis

Speech synthesis is used to produce synthetic speech from the text produces from different applications, an

978-1-4799-6393-5/14/$31.00 ©2014 IEEE

applet and the users. It is basically the technology if text to speech, there is following steps for producing text from the voice.

1) Structure analysis: In this we analyze that where the sentences and paragraphs starts and ends. it is also preferable for different punctuations and formatting of data.

2) Text pre-processing: This processing is special constructs of languages like English for their abbreviations, date, time, numbers, accounts and email address.

The remaining steps convert the spoken text to speech:

3) Text-to-phoneme conversion: In it we convert each words in the basic unit of sound in a language.

4) Prosody analysis: For determining the appropriate structure, words and prosody of the sentences.

5) Waveform production: For producing waveforms for each sentences by using prosody and phoneme information.

In the above steps there is the possibility of errors. The java speech API markup language uses to improve the output quality of the speech synthesizer.

B. Speech Recognition

This is basically used for determining the spoken language means what has been said. It converts speech in to the text.

It contains following steps:

1) Grammar design: It determines the words and patterns used in the spoken words.

2) Signal processing: It analyze the frequency characteristics of the incoming audio.

3) Phoneme recognition: Comparison of patterns between spectrums and phonemes.

4) Word recognition: It compares the phonemes in respect of words specified by active grammar.

5) Result generation: It provides the result of information about the words detected in incoming audio.

Grammar is one of the important part of speech reorganization because they contain the recognition process. These constraints provide more accuracy and more speedy.

Two basic grammar types supported by java speech api are rule grammars and dictation grammars. These both types are different from each other in various ways how result is provided, types of sentences it allow, set up of grammar in the applications, the amount of computational resources required and how the application design is used. JSAPI uses java speech grammar format for defining rule grammar.

Speech api are grouped by different packages. These packages contain class and interfaces.

The main three packages are:

javax.speech: For generic speech engine contains classes and interfaces.

javax.speech.synthesis: For speech synthesis contains classes and interfaces.

javax.speech.recognition: For speech recognition contains classes and interfaces.

All the applications of java speech api uses the engine manager class. This engine manager provides the static method for accessing speech recognition and synthesis.

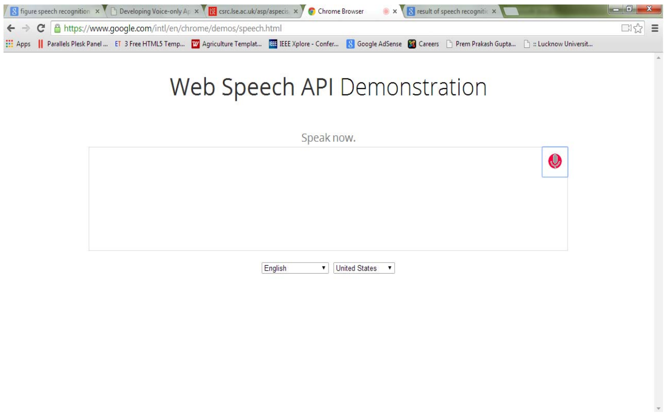

Figure.3.Demonstration of web speech

Speech applications use methods. These methods perform different actions like allocating and reallocating resources for speech engine, retrieving the properties and state of speech engine. The engine interfaces exposes the mechanism of pause and resuming the audio stream. Audio manager can be manipulated by engine interfaces.

The whole java speech api is work on event handling. Event generated can be easily identified and handled. Basically speech events handled through engine listener interface and also by synthesizer listener and recognizer listener.

V. CONCLUSION

The speech recognizer that we uses in our proposed work for making better interaction of web application with the help of voice is first open source. This recognizer is capable of real time, medium vocabulary speech recognition. By this we can recognize both words and number provide input in the form of voice and get the result in the form of voice. We basically runs it live mode but contains certain limitations. After all the limitations we tried to batch up the software and sound card input. This work can be taken is too far from the state in which it is now. The current software doesn't support voice to voice application on the web. This can be made more adaptable for any kind of web application. The author of this application is going to reduce the efficiency from the recognition.

Also, in the present work, we have hardcoded the using of web speech by voice to voice. This can be further programmed for making all web application is controlled by the help of voice. Besides all these, this application can be deployed for use on other handheld and mobile devices.

978-1-4799-6393-5/14/$31.00 ©2014 IEEE

REFERENCES

[1] L. R. Bahl, “Some experiments with large-vocabulary isolated-word sentence recognition”, Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84. Vol.9 1984.

[2] L. Rabiner and B.-H. Juang *Fundamentals of Speech Recognition*, Pearson Education, 2003.

[3] H. Palazç, “*A DCOM-Based Turkish Speech Recognition System: TREN – Turkish Recognition Engine*”, Computer and Information Sciences – ISCIS:Springer Berlin Heidelberg, 2005.

[4] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in speech Recognition”, In Proceedings of the IEEE 77, 257—286, 1989.

[5] L. Bahl , P. Brown , P. de Sow and R. Mercer "Maximum mutual information estimation of hidden Markov model parameters for speech recognition", Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP \'86), vol. 1, pp.49 -52 1986.

[6] R. K. Aggarwal, and M. Dave, “Performance evaluation of sequentially combined heterogeneous feature streams for Hindi speech recognition system”, Telecommunication Systems, Vol. 52, Issue 3, pp. 1457-1466 2013.

[7] Kai-Fu Lee, Hsiao-Wuen Hon, and Raj Reddy, “An Overview of the SPHINX Speech Recognition System”, IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 38, no.1, pp. 35-45, 1990.

[8] B.-H. Juang and S. Katagiri "Discriminative learning for minimum error classification", IEEE Trans. Signal Process., vol. 40, no. 12, pp.3043 -3054 1992.

[9] D. Povey and P. Woodland "Minimum phone error and I-smoothing for improved discriminative training", Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP \'02), vol. 1, pp.105 -108 2002.

[10] A.Ganapathiraju , J. Hamaker and J. Picone "Hybrid SVM/HMM architectures for speech recognition", Proc. 6th Int. Conf. Spoken Lang. Process. (ICSLP \'00), pp.504 -507 2000.

[11] http://www.speech.be.philips.com/index.htm

[12] R. Nisimura, et al. “*Development of Web-Based Voice Interface to Identify Child Users Based on Automatic Speech Recognition*” Human-Computer Interaction. Users and Applications, Springer Berlin Heidelberg, 2011.

[13] S.Miyamori, et al.: Real world utterance collection using voice-enabled web system for child speaker identification. In: Proc. Oriental COCOSDA 2010, paper-id 46 2010.

[14] Nisimura, R., et al. “Public Speech-Oriented Guidance System with Adult and Child Discrimination Capability”. In: Proc. IEEE ICASSP, vol. 1, pp. 433–436 2004.

[15] Lawrence R. Rabiner, “*Telecommunications Applications of Speech Processing*”, Speech Recognition and Coding: Springer Berlin Heidelberg 1995.

[16] E.McDermott, et al. "Discriminative training for large-vocabulary speech recognition using minimum classification error", IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 1, pp.203 -223 2007.

978-1-4799-6393-5/14/$31.00 ©2014 IEEE